Isolation, cloning and expression analysis of cinnamyl alcohol dehydrogenase (CAD) involved in phenylpropanoid pathway of *Erianthus arundinaceus*, a wild relative of sugarcane

Lakshmi Kasirajan¹ · Rabisha Valiyaparambth¹ · Aathisivan Kubandiran² · Janani Velu³

Received: 11 October 2019 / Accepted: 25 November 2019 / Published online: 4 December 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

In this study we have cloned and characterized cinnamyl alcohol dehydrogenase (CAD) involved in phenylpropanoid pathway which can be utilized for biomass modification for improved saccharification efficiency. The full length gene CAD is of 4 kb containing four exons and three introns, among which the exon 1 and 2 of 88 and 116 bp were conserved with sorghum and *Miscanthus* CADs. The coding region of CAD was identified with 1098 bp open reading frame (ORF), for 365 amino acids. In the PROSITE analysis, a zinc-containing alcohol dehydrogenase signature (GHEVVGEVVEVGPEV) and an NADP-binding domain motif (GLGGLG) was identified, while the motif analysis showed unique signature sequence of “LEPYLA” at 258–264 aa which was absent in the CAD sequences of other crops. This sequence information on CAD from *Erianthus* a bioenergy crop might be useful for subsequent research on lignin engineering for improved biomass conversion and for unravelling the impact of lignin on cell wall mechanics.

Keywords Bioenergy crop · Lignocellulosic biomass · Lignin · Secondary cell wall · Saccharification efficiency

Introduction

Second-generation bioethanol is a renewable energy resource produced from lignocellulosic biomass with the potential to reduce fossil fuel dependency. *E. arundinaceus* the immediate progenitor of *S. officinarum* has been identified as the most efficient converter of solar energy into stored chemical

energy and biomass accumulation. *Erianthus* spp. is characterized by high biomass production, multiple pest resistance and tolerance to drought, water logging and salinity. Besides *E. arundinaceus* is also an important source for fibre, now being used as a substitute for wood pulp in the paper industry. In spite of the economic importance of *Erianthus* for bioethanol production, the current understanding about the genetic information on biomass traits is limited. The main reason for limited usage is its high degree of complexity, due to its mixed composition of cellulose, hemicelluloses and lignin, in which the lignin is the compound most resistant to degradation. To effectively utilize cellulose and hemicellulose from lignocellulosic raw materials, it is essential to release them from lignin bonds. Lignin is a polymer comprised of chemically distinct monolignols, the abundance of which can vary among species, among individuals, and even among cell types within an organism. Based on the evidences obtained from our previous transcriptome studies in a set of sugarcane genotypes differing for lignin content, cinnamyl alcohol dehydrogenase (CAD) was exclusively differentially expressed in the high fiber clones playing major roles in genetic regulation of lignin production (Kasirajan et al. 2018), therefore we have cloned and characterized CAD from an *Erianthus* clone IK 76-81. Cinnamyl alcohol

✉ Lakshmi Kasirajan
K.Lakshmi@icar.gov.in; lakshmimbb@gmail.com

Rabisha Valiyaparambth
rabishavp9@gmail.com

Aathisivan Kubandiran
aathi9193@gmail.com

Janani Velu
janani_velu@srmuniv.edu.in

¹ Genomics Laboratory, Division of Crop Improvement, ICAR-Sugarcane Breeding Institute, Coimbatore 641007, India

² Department of Biotechnology, Rathnavel Subramaniam College of Arts and Science, Coimbatore 641402, India

³ Department of Biotechnology, SRM Institute of Sciences and Technology, Kattankulathur 603203, India

dehydrogenase (CAD, EC 1.1.1.195) catalyzes the conversion of the p-coumaric aldehyde, coniferyl aldehyde and sinapyl aldehyde to their corresponding alcohol derivatives; which is the last step in the lignin biosynthesis pathway. (Ma 2010; Preisner et al. 2014).

Materials and method

Plant material

Erianthus sp. clone IK 76-81, a wild relative of sugarcane was planted in 6-inche pots at Sugarcane Breeding Institute, Coimbatore. Six-month-old plants were used for the cloning CAD genes.

Total RNA isolation and cDNA conversion

Total RNA was isolated from stem tissue. About 2–2.5 g of stem tissue was ground to a fine powder in liquid nitrogen using pestle and mortar. The pulverized tissue was transferred to a 50-ml tube and homogenized with 5 ml of Trizol (Life Technologies) per gram of tissue, according to the manufacturer's instructions. RNA pellet was resuspended in 20 µl of warm diethyl pyro carbonate-treated water, vortexed gently for about 15 min followed by treatment with RNase-free DNase to remove the DNA contamination. Finally RNA sample was loaded on 1.0% agarose gel for quality check. First strand cDNA synthesis was performed with an oligodT primer and M-MuLV reverse transcriptase enzyme (cDNA kit, Thermo scientific) which was used as template for PCR amplification.

Primer designing

To clone the gene and the coding sequence of CAD, heterologous primers were designed using sequence information available in the public database from sugarcane and related plant species like sorghum, maize, and switch grass. Primers were designed flanking the genomic region as GEN_CAD and the coding region as CDS_CAD. Primer designing was done in such a way that its length is 20–27 nucleotide with 50–60% GC content. The primers used for expression studies are shown in Table 1.

Cloning of *EaCAD* gene and coding sequence

To clone the gene *cad*, GENCAD forward and reverse primer (Table 1) with the genomic DNA of *Erianthus* IK 76-81 was used as template and the PCR was carried out according to the following program: 94 °C for 3 min,

Table 1 Primer sequences used for gene and coding sequence cloning

S. no	Name of the primer	Sequence 5' to 3'
1	GENCAD forward	AACTCCCAGGAAGAGAATGGGGA
2	GENCAD reverse	ATCCATCAGCTCATCTGAAAGCCC
3	CDSCAD forward	ATGGGGAGCCTGGCGTCCGA
4	CDSCAD reverse	TCAGTTGCTCGGCGCATCG

followed by 35 cycles, 94 °C for 30 s, 61 °C for 30 s, 72 °C for 2 min, with a final extension step at 72 °C for 10 min with LA-TAQ, clontech TaKaRa a high fidelity polymerase enzyme. The PCR products were separated through a 1% agarose gel and eluted with gel elution columns (Qiagen, USA). The products fragments were cloned into pTZ57-R T/A cloning vector (Thermo scientific, USA) and sequenced by Department of Biochemistry, University of Delhi south campus, New Delhi using primer walking strategy subjected to restriction analysis and a map was generated (Fig. 1a, b).

Bioinformatics analysis

Nucleotide and amino acid sequences were analyzed using NCBI and EXPASY tools. Alignment of amino acid sequences was made with the Clustal X program, while BLAST searches were carried out using the National Center for Biotechnology Information Website (<http://www.ncbi.nlm.nih.gov>). Phylogenetic analysis was carried out using MEGA 4.0 program. The motif identification for these genes was detected using two independent programs, namely, PROSITE (<http://prosite.expasy.org/>) and CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The tertiary structures of CAD were predicted using Phyre2 software (<http://www.sbg.bio.ic.ac.uk/phyre2>). The Jmol (<http://www.jmol.org/>) program was used for the graphical representation of tertiary protein structure. The functional protein association networks were predicted with STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) <https://string-db.org/>. The MEME suite was used for predicting motifs on the EaCAD amino acid sequences (<http://meme-suite.org/tools/meme>). The parameters were set as follows: zero or one occurrence of a single motif per sequence; 15 as the maximum number of motifs to find. All other parameters were set at default.

Results and discussion

Cinnamyl alcohol dehydrogenase (CAD) catalyzes the reduction of cinnamaldehyde to p-coumaryl, coniferyl, and sinapyl alcohols during the last step of lignin biosynthesis.

Fig. 1 **a** Amplification of gene CAD from *Erianthus* with homologous primers. **b** Amplification of coding sequence of the gene CAD. **c** Map of the genomic region of CAD isolated from *Erianthus* showing the restriction pattern

Many researchers have identified and characterized many bonafide CAD from different plant species, however, no study has been reported on the *Erianthus* till now. The aim was to isolate, clone and characterize CAD gene from *Erianthus*.

Cloning of CAD from *Erianthus arundinaceus* clone IK 76-81

A full-length CAD gene of 4.0 kb in size was cloned from the *Erianthus* through homologues cloning strategy using LA-Taq DNA polymerase. Sequence analyses indicated that it contains four exons (exon1 of 88 bp, exon2 of 116 bp, exon3 of 219 bp and exon4 of 667 bp) and three introns. Generally CAD family is divided into three classes (Guo et al. 2010). All Class I CAD members, are bonafide CAD with identified 12 amino acids that constitute the substrate binding site of the angiosperms (Youn et al. 2006; Zhang et al. 2006; Bukh et al. 2012). In EaCAD, only two residues

were different; A60 and V95. The modified amino acids at these positions were similar to the results of CAD gene which was cloned from wheat by Ma (2010). According to Cheng et al. (2017) the identified EaCAD belong to the second type of CAD with three intronic and four exonic regions. An expected size of 1.1 kb was obtained while the total RNA was amplified with CDS_CAD forward and Reverse primers. The sequencing analysis confirmed an open reading frame (ORF) of 1098 bp encoding for 365 amino acid residues the results obtained were in accordance with other researchers (Fan and Li 2013; Deng et al. 2013; Tang et al. 2014). Using the BLAST search program (<http://www.ncbi.nlm.nih.gov/BLAST/>), EaCAD is the first reported full-length CAD from *Erianthus*. A phylogenetic analysis was performed to investigate the evolutionary relationships of the EaCAD gene with other plant species. Clustalw and phylogeny analysis of the EaCAD protein showed that it clustered with *Miscanthus*, *Panicum virgatum* the wild relatives of sugarcane and also identified bioenergy crops and later with sorghum,

Fig. 2 **a** Clustal W analysis of the amino acid sequence of CAD. **b** Phylogeny relationship of EaCAD with CAD from other crops

Zea mays and *S. officinarum* (Fig. 2a, b), while the motif analysis showed unique signature sequence “LEPYLA” at 258–264 which was absent in the CAD sequences of other crops like *Zea mays*, sorghum, switch grass and *S. officinarum* (Fig. 3a, b). As the identified EaCAD clustered with the other bonafide CAD enzymes from monocot and angiosperms, it belongs to group1. The calculated pI/Mw

of EaCAD is 6.06/38698.53 da. EaCAD consisted of secondary structure: 27% α -helixes, 25% β -sheet, 46% random coil regions. Most of the CAD reported earlier has a pI values from 5.33 to 8.77, while the length of amino acid sequences varied between 200 amino acids to 458 amino acids with a MWs of 22.3–49.1 kDa (Cheng et al. 2017).

(a)

(b) E-value: 6.1e+012 Site Count: 5 Width: 6

Name	Start	p-value	Sites
4. SoffCAD	198	5.03e-5	GLGGVGHMGV KVAKAM GHVTVISSS
2. SorCAD	196	9.70e-5	IVGLGGVGHM GVKVAK AMGHVTVIS
5. Zea	347	4.92e-4	LERNDVRYRF VVDVAG SNVEAEAAAA
3. MISCAD	216	1.21e-3	VTVISSSSSK RAEAMD HLGADAYLVS
1. ERICAD	258	7.80e-3	IIDTVPVHHP LEPYLA LLKLDGKHVL

Fig. 3 a Conserved motifs identified from the amino acid sequences of EaCAD. b The unique signature sequence identified in the EaCAD gene

Sequence analysis and three-dimensional structure prediction of EaCAD protein

The sequence of the gene obtained from *Erianthus* was further studied using NCBI conserved domain database the result showed specific hits GHE(X)2G(X)5G(X)2V motif involved in the Zn catalytic binding, the GD(X)10C(X)2C(X)2(X)7C structural Zn binding domain, and the conserved glycine-rich repeat G(X)GGV(L)G involved in NADP(H) binding site, as well as to Enoyl reductase PolyKetide Synthase super family (Fig. 4a). The identified motifs were in accordance with earlier report except an additional motif for Enoyl reductase PolyKetide Synthase super family (Eom et al. 2016; Xu et al. 2009). To understand the relationship between the *Erianthus* and sorghum CAD family members a comparative domain analysis was performed, which showed that the *Erianthus* CAD is 95%

similar to SbCAD2 while it is different from the other isoforms of CAD by domains distribution and position throughout the protein sequences (Fig. 4b).

The three-dimensional structure of EaCAD was predicted with RCSB PDB (<http://www.rcsb.org/pdb>). It showed a medium chain alcohol dehydrogenases catalytic domain with 195 residues in two fragments and a NAD(P) binding Rossmann domain with 157 residues. The results showed that two possible binding sites for Zn, which are being highlighted in the 3D model as ligand binding sites (Fig. 5a, b). The protein–protein interaction networks are important for system-level understanding and such predicted networks can provide novel targets for efficient interaction mapping. Therefore, the protein–protein interactions for CAD protein were predicted with STRING, which showed strong interactions with sorghum and Maize CAD with seven nodes with fifteen edges (Fig. 6). To understand the magnitude

(a)

Conserved domains on [lcl|Query_122507]

View Standard Results ?

ERICAD

(b)

Fig. 4 **a** Conserved domains identified from the amino acid sequences of EaCAD. **b** Prosite analysis of the EaCAD compared with sorghum CAD isoforms

of natural selection acting on EACAD protein we calculated the ratio of non-synonymous (K_a) to synonymous (K_s) substitution rates through Kimuras-two parameter (Yu et al. 2017). EaCAD had a ratio of K_a/K_s : 0.857, which was ($K_a/K_s < 1$), indicating that they have experienced purifying selection. Hence this sequence information on EaCAD can be used for downregulation of lignin through CRISPR/

CAS approach in Erianthus clones for improved saccharification efficiency. Therefore, to conclude this gene EaCAD repression in bioenergy crops might be useful for sinapaldehyde reduction during lignification of secondary cell wall, thereby to improve lignin digestibility for improved biomass conversion.

Fig. 5 **a** Secondary structure of EaCAD showing protein modification, dictionary of secondary structure of proteins (DSSP) sites. **b** The predicted 3D structure of EaCAD showing ligand-binding sites

Fig. 6 Network analysis of EaCAD with five green lines for neighborhood evidence with sorghum CAD and one blue line evidence for the database evidence

Acknowledgements This work has been supported by Indian Council of Agricultural Research (Grant no. P1-15/1-4). The authors wish to thank Dr. G. Hemaprabha, Head, Division of Crop Improvement, and Dr. A. Selvi, Principle Scientist, Biotechnology Section and Head, Sugarcane Breeding Institute, Coimbatore for kindly providing sugarcane clones and extending laboratory facility for carrying out the work, respectively.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by LK. The first draft of the manuscript was written by LK and all authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interests.

Data Archiving Statement All RNA-Seq read data has been submitted as sequence read archive (SRA) in NCBI, as mentioned in Kasirajan et al. (2018). Other relevant data are within the paper. The gene sequence of EaCAD cloned here will for public after the publication.

References

Bukh C, Nord-Larsen PH, Rasmussen SK (2012) Phylogeny and structure of the cinnamyl alcohol dehydrogenase gene family in *Brachypodium distachyon*. *J Exp Bot* 63(17):6223–6236

- Cheng X, Li M, Li D, Zhang J, Jin Q, Sheng L, Cai Y, Lin Y (2017) Characterization and analysis of CCR and CAD gene families at the whole-genome level for lignin synthesis of stone cells in pear (*Pyrus bretschneideri*) fruit. *Biol Open* 6(11):1602–1613
- Deng WW, Zhang M, Wu JQ, Jiang ZZ, Tang L, Li YY, Wei CL, Jiang CJ, Wan XC (2013) Molecular cloning, functional analysis of three cinnamyl alcohol dehydrogenase (CAD) genes in the leaves of tea plant, *Camellia sinensis*. *J Plant Physiol* 170(3):272–282
- Eom SH, Kim H, Hyun TK (2016) The cinnamyl alcohol dehydrogenase (CAD) gene family in flax (*Linum Usitatissimum* L.): insight from expression profiling of CADs induced by elicitors in cultured flax cells. *Arch Biol Sci* 68(3):603–612
- Fan B, Li F (2013) Cloning and analysis of gene fragments encoding C4H and CAD from *Miscanthus sinensis*. *Plant Gene Trait* 4(6):30–36
- Guo DM, Ran JH, Wang XQ (2010) Evolution of the cinnamyl/sinapyl alcohol dehydrogenase (CAD/SAD) gene family: the emergence of real lignin is associated with the origin of Bona Fide CAD. *J Mol Evol* 71:202–218
- Kasirajan L, Hoang NV, Furtado A, Botha FC, Henry RJ (2018) Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content. *Sci Rep* 8(1):11612
- Ma QH (2010) Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *J Exp Bot* 61(10):2735–2744
- Preisner M, Kulma A, Zebrowski J et al (2014) Manipulating cinnamyl alcohol dehydrogenase (CAD) expression in flax affects fibre composition and properties. *BMC Plant Biol* 14:50
- Tang R, Zhang X, You-Han L, Xie X (2014) Cloning and in silico analysis of a cinnamyl alcohol dehydrogenase gene in *Pennisetum purpureum*. *J Genet* 93:145–158
- Xu Z, Zhang D, Hu J et al (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics* 10:S3. <https://doi.org/10.1186/1471-2105-10-S11-S3>
- Youn B, Camacho R, Moinuddin SG, Lee C, Davin LB, Lewis NG, Kang C (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org Biomol Chem* 4:1687–1697
- Yu Y, Liu J, Feng N, Song B, Zheng Z (2017) Combining sequence and gene ontology for protein module detection in the weighted network. *J Theor Biol* 412:107–112
- Zhang K, Qian Q, Huang Z, Wang Y, Li M, Hong L, Zeng D, Gu M, Chu C, Cheng Z (2006) GOLD HULL AND INTERNODE2 encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice. *Plant Physiol* 140(3):972–983